Identity of the RNA-binding protein K of hnRNP particles with protein H16, a sequence-specific single strand DNA-binding protein

Claire Gaillard*, Eric Cabannes and François Strauss
Institut Jacques Monod, 2 place Jussieu, 75251 Paris 05, France

ABSTRACT

Protein H16, which we have identified previously in mammalian cell lines, binds *in vitro* to two single stranded DNA sites on the late strand of the early promoter of SV40. It has no other single strand binding site in the SV40 genome and does not bind to double stranded DNA. *In vitro*, H16 can be shown to stimulate strongly the activity of purified RNA polymerase II. Here we have purified this 70 kDa protein from cultured monkey cells and have sequenced three of its tryptic peptides. The analysis indicates that H16 is the simian homolog of human protein K, a nuclear RNA-binding protein found in heterogeneous nuclear ribonucleo-protein (hnRNP) particles, which contains a KH domain present in several proteins including the fragile X mental retardation gene product (FMR1). The binding affinities of protein K/H16 for RNA and DNA were subsequently compared in detail. They showed that under conditions where K/H16 binds strongly to its single stranded DNA site, it binds very weakly to the corresponding RNA sequence. This result suggests a possible shuttling of the protein from RNA to DNA during processes which involve opening of the DNA double helix.

INTRODUCTION

Whereas most known DNA-binding proteins bind to specific sites on double-stranded DNA, several others have been found which bind instead to sites on single-stranded DNA. Such sequence-specific single strand binding proteins might possibly play a regulatory role during replication and transcription, as both of these processes require opening of the DNA double helix. Single stranded DNA, in fact, presents a much wider conformational variability than double stranded DNA and could thus form an excellent substrate for highly specific interactions with regulatory proteins.

One of the first single strand DNA binding proteins to be studied was the protein H16 (1). It was identified in nuclear extracts from cultured monkey cells due to its capacity for interacting with the control region of the SV40 genome *in vitro*.

H16 was shown to have two C-rich binding sites on the late strand of the early SV40 promoter, in the region of the 21 base pair (bp) repeats. No binding could be detected to early strand DNA, nor to the double stranded DNA, nor to RNA transcripts of this region (1). It was later observed that protein H16 had no additional binding site in the SV40 genome and that it could stimulate the activity of purified RNA polymerase II *in vitro* (2). Its apparent molecular weight was found to be 70 kDa in SDS (2).

In the present work, protein H16 has been purified and the amino acid sequences of three of its tryptic peptides determined. This analysis showed protein H16 to be analogous to human protein K, a protein associated with primary transcripts of RNA polymerase II in hnRNP particles (3, 4; see reference 5 for a review). Protein K is the major poly(C)-binding protein in HeLa cells, and it has been shown by immunofluorescence to be localized in the cell nucleus (6, 7). The amino acid sequence of protein K contains a motif, the KH motif, which shows significant homology to several proteins some of which are known nucleic acids binding proteins (8, 9). For example, the product of the fragile X mental retardation gene FMR1 contains a KH domain (10, 11) which is essential for RNA binding *in vitro* (12). Protein K has also recently been shown to bind protein Src (13). Since protein K was initially identified as an RNA binding protein (4, 6) and H16 as a single strand DNA binding protein (1, 2), we have also analyzed here the relative binding properties of protein H16 for DNA and RNA. Results show that under conditions of strong binding to its single stranded DNA site, protein H16 binds only very weakly to the corresponding RNA sequence. These observations suggest an hypothesis concerning the possible function of protein K/H16 *in vivo*.

MATERIALS AND METHODS

Purification of protein H16

Extraction of nuclear proteins. Culture of African green monkey kidney cells (CV1 line) and purification of nuclei were as described (1, 2). Nuclei were prepared from approximately 5×10^9 cells and resuspended in 0.4 M NaCl, 25 mM Tris−HCl pH 7.5, 10 mM K phosphate pH 7.5, 2 mM dithiothreitol, containing four protease inhibitors (Boehringer): 2 µg/ml

pepstatin, 10 μg/ml bestatin and chymostatin, and 25 μg/ml leupeptin. After 30 min at 0°C with occasional gentle agitation, followed by centrifugation at 10,000×g for 30 min, glycerol was added to the supernatant to a final concentration of 15%.

Hydroxyapatite chromatography. Two characteristics of hydroxyapatite were useful for the purification of protein H16. First, H16 and many other nuclear proteins bind to hydroxyapatite in the presence of high concentrations of NaCl, allowing nuclear extracts to be loaded on the column with no previous dialysis step. Second, H16 is one of the few proteins, thus bound, to elute with a KCl concentration gradient, while most other proteins remain bound, eluting only with a gradient of K phosphate. Therefore, after direct loading of nuclear extracts onto a column of 1.6×20 cm, the column was washed with 0.4 M NaCl, 10 mM K phosphate pH 7.5, 1 mM DTT, 15% glycerol. Protein H16 was eluted with 0.4 M KCl, 10 mM K phosphate pH 7.5, 1 mM DTT, 15% glycerol. The presence of protein H16 in the fractions was assayed by gel retardation.

HPLC-Mono Q chromatography. Fractions from the hydroxyapatite column containing protein H16 were pooled, dialyzed against 30 mM NaCl, 25 mM Tris—HCl pH 7.5, 1 mM DTT, 15% glycerol, loaded on a Mono Q column (Pharmacia), and eluted with a linear NaCl concentration gradient, from 30 mM to 0.6 M, in 25 mM Tris—HCl pH 7.5, 1 mM DTT, 15% glycerol. Protein H16 eluted as a sharp peak at 250 mM NaCl.

Peptide cleavage and sequencing

The entire H16 preparation of approximately 40 μg was submitted to electrophoresis on a 10% polyacrylamide—SDS gel and then transferred to an Immobilon membrane. Following rapid staining with amidoblack, the band containing H16 was cut out, digested with trypsin, and peptides fractionated by HPLC. Three peaks were selected and sequenced using an Applied Biosystems 470A Protein Microsequencer.

Interaction of protein H16 with DNA and RNA

Binding assay. Incubation of H16 with nucleic acids was performed as described previously (1, 2), except for the systematic addition of RNAse inhibitor (RNAsin from Promega) at 0.3 units/μl. The complexes formed were analyzed by gel retardation assay and autoradiography (1, 2). When necessary, radioactivity in the bands was quantitated using a Phosphorimager (Molecular Dynamics).

Labelled DNA and RNA. The late strand of the StyI fragment from the control region of SV40 was purified and ³²P end-labelled using polynucleotide kinase, as described (1). Transcripts of the SV40 control region were synthesized *in vitro* using T7 RNA polymerase and constructs described previously which contain this sequence cloned in either orientation adjacent to a T7 promoter (1). The RNA transcripts were dephosphorylated with calf intestine phosphatase (Boehringer), ³²P end-labelled using polynucleotide kinase (New England Biolabs), and purified by gel electrophoresis and electroelution.

Non-radioactive DNA and RNA. Single-stranded DNA used as competitor in the binding assays was *E.coli* DNA sonicated to an average length of 1 kbp and heat denatured. Poly(C) and poly(dC) were from Pharmacia. Total rat liver RNA and rat liver polyA⁺ RNA were from Stratagene.

RESULTS AND DISCUSSION

Protein H16 was purified by column chromatography of nuclear extracts from cultured monkey cells, using the gel retardation assay to monitor its presence in the fractions by its specific binding to the SV40 early promoter. Using a known amount of labelled DNA as aprobe, measurement of the amount of shifted DNA in this assay gave an estimate of the amount of active protein present in the fractions. After chromatography on hydroxyapatite and HPLC on Mono-Q, we obtained a purified protein fraction containing 30 μg of protein H16 with specific binding activity. Preparative SDS—polyacrylamide gel electrophoresis of the whole preparation and transfer to an Immobilon membrane yielded a prominent band containing about 40 μg of total protein, showing that no more than 25% of the protein in the band was comprised of H16 molecules that had lost their DNA-binding activity during the purification or of minor contaminants. The band was subsequently digested with trypsin, peptides were fractionated by HPL, and three well-separated, major peaks were selected for sequencing. This ensured that the sequence data did not correspond to a minor contaminant in the H16 preparation.

A search of the GenBank database for homology with these sequences revealed that all three were found within the amino-acid sequence of protein K, a component of hnRNP from human cells. No other significant homology with any other sequence in the databases was found. Figure 1 shows the amino-acid sequence of protein K, as deduced from the nucleotide sequence of its cloned cDNA (6), andthe sequences of the H16 peptides. The first peptide shows 13 amino acids identical to those found in protein K, while 3 additional amino acids could not be unambiguously determined from the sequencing data. The 12 residues of the second peptide show a complete identity with those of protein K. The third peptide shows 11 of 13 amino acids to be identical with those of protein K, with two differences due perhaps to the different species origins of proteins K and H16, human and simian, respectively.

In addition to these striking sequence homologies, proteins H16 and K share other properties. First, their have identical molecular weights in SDS, 70 kDa (2, 6), and identical isoelectric points, 5.5 (7, and data not shown). Second, both H16 and K bind strongly to poly(dC) and not to the other three DNA homopolymers (2, 4, 6, 7). Third, we have verified that H16 has the same affinity for poly(C) as for poly(dC) (data not shown), which is another known property of protein K (6).

METEQPEETF	PNTETNGEFG	KRPAEDMEEE	QAFKRSRNTD	EMVELRILLQ	50
SKNAGAVIGK	GGKNIKALRT	DYNASVSVPD	SSGPERILSI	SADIETIGEI	100
	T	DYNASVXVPD	XXGPX		
LKKIIPTLEE	GLQLPSPTAT	SQLPLESDAV	ECLNYQHYKG	SDFDCELRLL	150
IHQSLAGGII	GVKGAKIKEL	RENTQTTIKL	FQECCPHSTD	RVVLIGGKPD	200
RVVECIKIIL	DLISESPIKG	RAQPYDPNFY	DETYDYGGFT	MMFDDRRGRP	250
VGFPMRGRGG	FDRMPPGRGG	RPMPPSRRDY	DDMSPRRGPP	PPPPGRGGRG	300
GSRARNLPLP	PPPPPRGGDL	MAYDRRGRPG	DRYDGMVGFS	ADETWDSAID	350
TWSPSEWQMA	YEPQGGSGYD	YSYAGGRGSY	GDLGGPIITT	QVTIPKDLAG	400
			GSY	GDLGGPIIT	
SIIGKGGQRI	KQIRHESGAS	IKIDEPLEGS	EDRIITITGT	QDQIQNAQYL	450
		IDEPLEGS	EDAFI		
LQNSVKQYSG	KFF	463			

Figure 1. Amino acid sequence of protein K and of three tryptic peptides from protein H16.

The data thus show that protein H16 and protein K are homologous proteins from two different species. All the data available on the binding of protein H16 to single stranded DNA and of protein K to RNA and hnRNP particles can now be considered to regard the same protein, denoted K/H16.

This conclusion raises questions concerning the relative affinities of K/H16 for RNA and single stranded DNA. Since in our early work we were unable to detect the formation of a complex between protein H16 and the RNA transcripts of the control region of SV40 (1), the following experiments were thus performed to better define the binding characteristics of K/H16.

First, the binding of H16 to its single strand DNA site on the late strand of SV40 early promoter was studied in the presence of different non-radioactive competitors (Figure 2). It may be observed that single strand DNA from *E.coli* and polyA$^+$ RNA compete with the same efficiency, which is weak but significant (lanes 4−11). As expected, non-radioactive late strand DNA, which contains the H16 binding sites, competes very strongly (lanes 2−3). The absence of visible competition from total RNA is probably due to its high content of RNA with double stranded secondary structures (tRNA, rRNA). Therefore, the affinity of H16 for mixed sequence RNA is similar to its affinity for single stranded *E.coli* DNA, which is much weaker than its affinity for its site in the SV40 early promoter. Quantitation of these data shows that *E.coli* single stranded DNA and polyA$^+$ RNA are about 150 times less efficient as competitors than the late strand DNA. As the fragment of the late strand is 224 nucleotides long and contains two binding sites for H16, this can be interpreted in terms of the ratio of specific to non-specific affinities. Either the non-specific affinity of H16 for random DNA or RNA sites is 1.6×10^4 times lower than its affinity for its specific DNA sites on the promoter, or *E.coli* DNA and polyA$^+$ RNA contain, on average, strong binding sites only every 16,000 nucleotides. The actual situation lies probably somewhere between these two extremes.
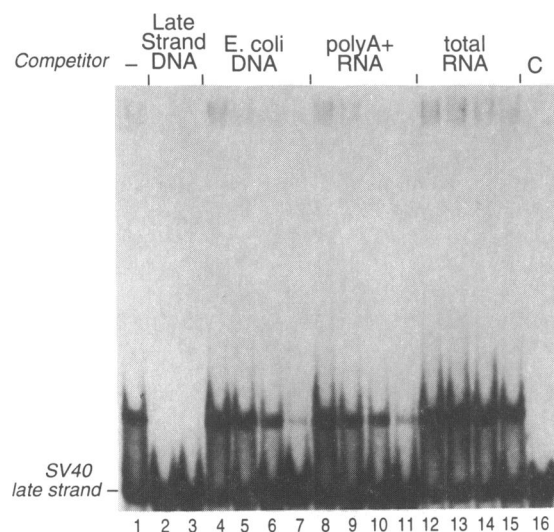
In the following experiment, the interaction of H16 with labelled RNA transcripts of defined sequence was studied (Figure 3). Both transcripts from the SV40 control region, synthesized *in vitro* using T7 RNA polymerase, were ^{32}P end-labelled and used in gel retardation experiments in parallel with the late strand DNA of the early promoter. Variable amounts of protein were used. At the higher protein concentration, two retarded bands are observed with DNA, corresponding to the two H16 sites on DNA. Under the same conditions, no retarded band appears with the RNA of complementary sequence, and two retarded bands are only barely visible with the RNA of homologous sequence. It can thus be concluded that, while H16 binds strongly to specific sites on late strand DNA, it binds extremely weakly to the RNA of corresponding sequence. This result is reminiscent of the work on DNA and RNA aptamers, which showed several examples of single-stranded DNAs interacting with specific ligands while RNAs of identical sequences could not interact with the same ligands (14). This does not exclude the possibility that strong binding sites for K/H16 exist elsewhere on RNA. K/H16 binds very strongly to poly(C) and weakly to the late strand RNA of the early promoter, which contains 60% C in the 21 bp repeats where the H16 sites are located. Therefore, RNA sequences with a high cytosine content would almost certainly bind K/H16 strongly.

Protein K has recently been shown to be heterogeneous and composed of four variants as a result of differential RNA splicing (7). However, isoelectric focusing of our purified H16 preparation shows a single band, with a pI of 5.5 corresponding to variant A of protein K. Similarly, protein K extracted from purified hnRNP particles contains only variant A (7). This has been suggested to reflect the differential binding of the variants to particles in nuclei (7). An alternative possibility could be that variants have different affinities for nucleic acids, although the sequence variations are all located outside of the KH domain, which has been identified as essential for binding of protein K to poly(dC) and to RNA (12).
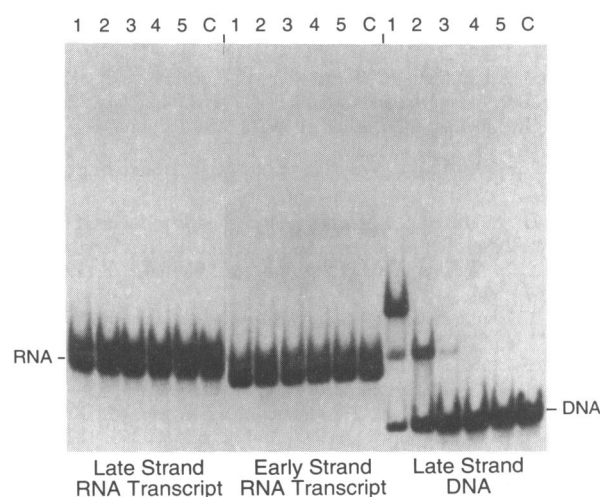

Figure 2. Interaction of protein H16 with the late strand DNA of SV40 in the presence of different non-radioactive competitors. Lane 1: no competitor; lanes 2−3: unlabelled late strand DNA, 1 ng and 5 ng; lanes 4−7: heat-denatured *E.coli* DNA, 2, 8, 30 and 125 ng, respectively; lanes 8−11: polyA$^+$ rat liver RNA, 2, 8, 30 and 125 ng; lanes 12−15: total rat RNA, 2, 8, 30 and 125 ng; lane 16: control, no protein added.

Figure 3. Interaction of protein H16 with labelled RNA transcripts from the SV40 control region. Transcripts were synthesized *in vitro* with T7 RNA polymerase using either DNA strand as template. They were ^{32}P end-labelled and used in a gel retardation experiment, in parallel with late strand DNA, with three-fold serial dilutions of protein H16 in lanes 1−5. Lanes C: controls, no protein added. No unlabelled competitor was used in this experiment.

Two hnRNP proteins, K/H16 described here and A2/B1, are now known to possess specific binding sites on single stranded DNA. Protein K/H16 has a specific site on the late strand of SV40 early promoter (1, 2) and in the promoter of human c-*myc* gene (15). The rat homolog of K/H16 has been shown to bind specifically to a single stranded DNA site in the rat albumin gene (16). A2/B1 has also been shown to possess specific binding sites on single-stranded DNA (17). It will be interesting to study the specificity of interaction of other hnRNP proteins with single stranded DNA.

It will also be of interest to elucidate the possible role of sequence-specific single strand DNA binding proteins on DNA processes which require opening of the double helix, such as replication and transcription. Stimulation of RNA polymerase II activity *in vitro* by protein H16 (2) and stimulation of c-*myc* gene expression by cotransfection with a vector expressing protein K (15) strongly suggest that K/H16 could play a role in transcription. Protein K/16 might, for example, shuttle between sites of low affinity on RNA and a few high affinity binding-sites on single stranded DNA which would be accessible only during the opening of the DNA double helix.

ACKNOWLEDGEMENTS

REFERENCES

1. Gaillard, C., Weber, M. and Strauss, F. (1988) *J. Virol.*, **62**, 2380−2385.
2. Gaillard, C. and Strauss, F. (1990) *J. Mol. Biol.*, **215**, 245−255.
3. Piñol-Roma, S., Choi, Y.D., Matunis, M.J. and Dreyfuss, G. (1988) *Genes Dev.*, **2**, 215−227.
4. Swanson, M.S. and Dreyfuss, G. (1988) *Mol. Cell. Biol.*, **8**, 2237−2241.
5. Dreyfuss, G., Matunis, M.J., Piñol-Roma, S. and Burd, C. (1993) *Annu. Rev. Biochem.*, **62**, 289−321.
6. Matunis, M.J., Michael, W.M. and Dreyfuss, G. (1992) *Mol. Cell. Biol.*, **12**, 164−171.
7. Dejgaard, K., Leffers, H., Rasmussen, H.H., Madsen, P., Kruse, T.A., Gesser, B., Nielsen, H. and Celis, J.E. (1994) *J. Mol. Biol.*, **236**, 33−48.
8. Siomi, H., Matunis, M.J., Michael, W.M. and Dreyfuss, G. (1993) *Nucl. Acids Res.*, **21**, 1193−1198.
9. Gibson, T.J., Thompson, J.D. and Heringa, J. (1993) *FEBS Lett.*, **324**, 361−366.
10. Siomi, H., Siomi, M.C., Nussbaum, R.L. and Dreyfuss, G. (1993) *Cell*, **74**, 291−298.
11. Ashley, C.T., Wilkinson, K.D., Reines, D. and Warren, S.T. (1993) *Science*, **262**, 563−566.
12. Siomi, H., Choi, M., Siomi, M.C., Nussbaum, R.L. and Dreyfuss, G. (1994) *Cell*, **77**, 33−39.
13. Taylor, S.J. and Shalloway, D. (1994) *Nature*, **368**, 867−871.
14. Ellington, A.D. and Szostak, J.W. (1992) *Nature*, **355**, 850−852.
15. Takimoto, M., Tomonaga, T., Matunis, M., Avigan, M., Krutzsch, H., Dreyfuss, G. and Levens, D. (1993) *J. Biol. Chem.*, **268**, 18249−18258.
16. Flavin, M. and Strauss, F. (1991) *DNA and Cell Biol.*, **10**, 113−118.
17. McKay, S.J. and Cooke, H. (1992) *Nucl. Acids Res.*, **20**, 6461−6464.